Risk of an estimator

Utility functions

\[ U(\theta, \hat{\theta}) \] loss
utility derived from estimating \( \theta \) using \( \hat{\theta} \).

Loss functions: What do we want them to look like?

\[ L(\theta, \theta) = 0 \]
\[ L(\theta, \hat{\theta}) \geq 0 \]
\[ L(\theta, \hat{\theta}) \] to be nondecreasing as \( \hat{\theta} \) moves away from \( \theta \).

Standard choices

\[ L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2 \] symmetric
\[ L(\theta, \hat{\theta}) = |\theta - \hat{\theta}| \]
$R(\theta) = \text{risk of an estimator}$

$R(\theta) = E(L(\theta, \hat{\theta}))$

$= \int L(\theta, \hat{\theta}) f(\hat{\theta}) \, d\hat{\theta}$

$= \int L(\theta, \hat{\theta}) f(x_1, \ldots, x_n) \, dx_1 \ldots dx_n$

because $\hat{\theta} = \hat{\theta}(x_1, \ldots, x_n)$

\textbf{Example:}

$x_1, x_2, \ldots, x_{25} \sim N(\theta, 1)$

$L(\theta, \hat{\theta}) = (\theta - \hat{\theta})^2$

$\hat{\theta}_1 = 0$

$\hat{\theta}_2 = \bar{x}$

$R_1(\theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta - 0)^2 f(x_1, \ldots, x_{25}) \, dx_1 \ldots dx_{25}$

$R_2(\theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (\theta - \bar{x})^2 f(x_1, \ldots, x_{25}) \, dx_1 \ldots dx_{25}$

$= \frac{\theta^2}{25}$
\[ \int \cdots \int (\theta^2 - 2\theta \bar{X} + \bar{X}^2) f(x_1, \ldots, x_n) \, dx_1 \cdots dx_n \]

\[ = \bar{\theta} - 2\theta \mathbf{E}(\bar{X}) + \mathbf{E}(\bar{X}^2) \]

\[ = \bar{\theta}^2 - 2\theta^2 + \text{Var}(\bar{X}) + [\mathbf{E}(\bar{X})]^2 \]

\[ = \theta^2 - 2\theta^2 + \text{Var}(\bar{X}) + \theta^2 = \text{Var}(\bar{X}) = \frac{1}{25} \]

\[ \text{It is hard to find estimators that uniformly dominate every other estimator.} \]

\[ \text{What can be done is restrict the class somehow (for example, to unbiased estimators) and find "optimal" estimators in that class.} \]
If you restrict attention to unbiased estimators and the loss function is assumed to be quadratic, then we recover the MVUE.

Likelihood principle

If two models lead to likelihoods that are proportional to each other, then the inferences derived from both models should be the same.

Experiment 1:
You toss a coin a fixed number of times (12) and you record the number of heads (5).

Experiment 2:
You toss the coin until you get exactly heads, and you happen to observe 12 tosses.
We want to infer the value of $p$, the probability of heads.

Exp 1: \[ L(p) = \binom{12}{5} p^5 (1-p)^7 \]

Exp 2: \[ L(p) = \binom{11}{4} p^5 (1-p)^7 \]

The likelihoods are the same, but they are proportional. A procedure that follows the likelihood principle should yield the same result when applied to both likelihoods.

Does MLE follow the lik. principle?

Yes

Does hypothesis testing follow the lik principle?

No! Not obvious

\[ \Rightarrow \text{Show it} \]

$H_0: p = p_0 \quad \text{vs} \quad H_a: p \neq p_0 \text{ (Exact test)}$
Can we find low dimensional summaries of the data that allow us to perform inferences without the original dataset?

\[ \text{Answer is model dependent!!} \]

We say that a statistic \( S \) is sufficient if \( S = S(x_1, \ldots, x_n) \) is independent of

\[
\frac{f(x_1, \ldots, x_n)}{f(S(x_1, \ldots, x_n))}
\]

Consider \( x_1, x_2, \ldots, x_n \sim \text{Poi}(\lambda) \)

Is \( \sum_{i=1}^{n} x_i \) a sufficient statistic in this distribution?
If \( X_1, \ldots, X_n \sim \text{Poi}(\lambda) \),
\[ S = \sum_{i=1}^{n} X_i \sim \text{Poi}(n\lambda) \]

\[
f(x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum x_i}}{(2\pi x_i)!} \]

\[
f(S(x_1, \ldots, x_n)) = \frac{e^{-n\lambda} \lambda^{\sum x_i}}{e^{-\lambda} \lambda^{\sum x_i} (\Pi x_i)!} = \frac{n^{\sum x_i} \lambda^{\sum x_i}}{(\Pi x_i)}
\]

Since the ratio does not depend on \( \lambda \), then this is a sufficient statistic.

**Factorization theorem**

If 
\[
f(x_1, \ldots, x_n | \theta) = f(x_1, \ldots, x_n | \theta) \]
\[= K_1(S(x_1, \ldots, x_n), \theta) K_2(x_1, \ldots, x_n) \]

then \( S(x_1, \ldots, x_n) \) is sufficient.
Examples:

\[ x_1, \ldots, x_n \sim \text{Unif} [\theta, \theta + 1] \]

\[ f(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } \theta \leq x_1 \leq \theta + 1 \\
0 & \text{otherwise} \\
\end{cases} \]

\[ f(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } \max \{x_i\} \leq \theta + 1 \\
0 & \text{otherwise} \\
\end{cases} \]

\[ f(x_1, \ldots, x_n) = \begin{cases} 1 & \text{if } \frac{\max \{x_i\} - 1}{\min \{x_i\}} \leq \theta \leq \min \{x_i\} \\
0 & \text{otherwise} \\
\end{cases} \]

Indicator functions

\[ f(x_1, \ldots, x_n) = \begin{cases} 1 & \theta \in [\max \{x_i\} - 1, \min \{x_i\}] \\
0 & \text{otherwise} \\
\end{cases} \]

\[ k_1(s, \theta) = \begin{cases} 1 & \theta \in [\max \{x_i\} - 1, \min \{x_i\}] \\
0 & \text{otherwise} \\
\end{cases} \]

\[ k_2(x_1, \ldots, x_n) = 1 \iff s = (\max \{x_i\}, \min \{x_i\}) \]
If \( S(x_1, \ldots, x_n) \) is a sufficient statistic, then \( g(S(x_1, \ldots, x_n)) \) is also a sufficient statistic.

**Minimal sufficient statistic**

The order statistic (for iid data) is always a sufficient statistic. But we can typically reduce even further.

A sufficient statistic is called minimal if for two different datasets \( x_1, \ldots, x_n \) and \( y_1, \ldots, y_n \) (of the same size) the ratio

\[
\frac{L(\theta; x_1, \ldots, x_n)}{L(\theta; y_1, \ldots, y_n)}
\]

is constant as a function of \( \theta \) if and only if

\[ S(x_1, \ldots, x_n) = S(y_1, \ldots, y_n) \]
In many cases the minimal sufficient statistic has the same dimension as the parameter space \( \Rightarrow \) Not always true!!

Another example:

Double exponential density:

\[
    f(x; \lambda, \mu) = \frac{1}{2\lambda} \exp\left\{-\frac{|x - \mu|}{\lambda}\right\}
\]

\(-\infty \leq x \leq \infty\)

The minimal sufficient statistic for this problem is the order statistics.

Rao-Blackwellization: how do we get MVUE? Let \( x_1, \ldots, x_n \) iid \( f(x; \theta) \) let

\( Y = S(x_1, \ldots, x_n) \) be sufficient for \( \theta \) and consider \( Z = g(x_1, \ldots, x_n) \) but not of \( Y \) alone, with \( Z \) being unbiased.
Let \( W = E[\xi | Y] = E[g(x_1, \ldots, x_n) | S(x_1, \ldots, x_n)] \). Then:

1) \( W \) is an unbiased estimator.
2) It has a lower variance than \( \xi \).
3) \( \xi \) is a function of the sufficient statistic.

\[
\text{Var}(\xi) = \frac{\text{Var}(E(\xi | Y))}{E\left[\text{Var}(\xi | Y)\right]}
\]

\[
E(\xi | Y) = E(E(\xi | Y))
\]

\[
\text{Var}(\xi) = \text{Var}(E(\xi | Y)) + E(\text{Var}(\xi | Y))
\]

\[
\Rightarrow \text{Var}(\xi) \geq \text{Var}(E(\xi | Y))
\]

- Relationship between MLEs and sufficient statistics.

If the MLE exists and is unique, then it is a function of the sufficient statistic.
Counterexample

\( X_1, \ldots, X_n \sim \text{Uni} \{0, \theta+1\} \)

Completeness

A family of distributions \( \{f(x|\theta) : \theta \in \Theta\} \) is said to be complete if

\[
\mathbb{E}(g(x)) = 0 \Rightarrow g(x) = 0 \text{ for all } x \text{ and } \theta,
\]

except for a set of points with measure zero.

\[
f(x|\theta) = \frac{e^{-\theta} \theta^x}{x!}
\]

\[
\mathbb{E}(g(x)) = \sum_{x=0}^{\infty} g(x) \frac{\theta^x}{x!} = e^{-\theta} \sum_{x=0}^{\infty} \frac{\theta^x}{x!} \mathbb{E}(g(x)) = e^{-\theta} \left( g(0) + \frac{g(1)}{1} + \frac{g(2)}{2} + \ldots \right)
\]

One possibility is

\( g(0) = 1, \quad g(1) = -\frac{1}{\theta}, \quad g(2) = \frac{2}{\theta^2}, \quad g(3) = -\frac{6}{\theta^3} \)

Does not work

The only option that works is \( g(0) = g(1) = \ldots = 0 \)
Lehmann and Scheffe's theorem

\( X_1, \ldots, X_n \) is an iid sample with pdf \( f(x_i | \theta) \) and \( Y = s(X_1, \ldots, X_n) \) is a sufficient statistic for \( f(x_i | \theta) \) and if the family for \( f(y | \theta) \) is complete and there is a function of \( Y \) that is an unbiased estimator of \( \theta \), then this function is the MVUE.

Let \( X_1, \ldots, X_n \sim \text{Exp}(\lambda) \). Find an unbiased estimator for \( \mu = \frac{1}{\lambda} \).

\[
L(x_1, \ldots, x_n | \lambda) = \left( \frac{1}{\lambda} \right)^n \exp \left\{ -\frac{\sum x_i}{\lambda} \right\} \]

\[= \sum x_i \text{ is sufficient.} \]

\[\sum x_i \sim \text{Gamma}(n, \lambda) \Rightarrow \]

Show that \{Gamma \( (n, \lambda) : \lambda > 0 \} is complete.

The MVUE has to be a function of \( \sum x_i \).

Start with \( \bar{X} = \frac{1}{n} \sum x_i \Rightarrow \text{E}(\bar{X}) = \frac{n-1}{n} \lambda \).